

What is claimed is:

1. A data management system, comprising:

a first server processor for restoring a plurality of received data files, the
5 data files being capable of being different file types;

a file organizing/categorizing processor for organizing the received data
files, based on a predetermined user list, into a source directory structure and a
destination directory structure;

a file logging processor for logging the received data files into a database
10 formed by the source and destination directory structures and identifying a file
type of the received data files;

a de-duplicate processor for calculating a SHA value of the received data
files to determine whether the received data files have duplicates and flagging
duplicated data files in the database;

15 an image conversion processor for converting the remaining, de-
duplicated, data files into image files, respectively; and

a second server processor for exporting the image files.

2. The system of claim 1, wherein the image files are stored in the database to be
20 viewed.

3. The system of claim 1, wherein the image files converted from the data files are in
a tiff format.

4. The system of claim 1, wherein the data files include email data files and user data files.

5. The system of claim 4, wherein the email data files are in a variety of formats including Microsoft Mail, Outlook, GroupWise, Lotus Notes, , the user data files have a variety of formats including Word, Excel, PowerPoint, and Access.

6. The system of claim 4, wherein the email data files include attachment data and email files.

7. The system of claim 6, wherein the attachment data and email files are associated with the email data files such that the image data files for the email data files and the corresponding attachment data and email files can be viewed together.

8. The system of claim 1, wherein the file logging processor, the image conversion processor, and the second server processor are parallel processors such that the data files are parallel-processed in a data file logging stage, an image conversion stage, and an image file output stage.

9. The system of claim 1, wherein the data files having the same file type are converted into the image files together.

10. The system of claim 1, wherein the data management system includes a plurality of image conversion processors, each of the image conversion processors being capable of converting the data files having the same file type into the corresponding image files.

5 11. The system of claim 1, wherein the file logging processor identifies the file type of the data files based on the SHA value and a file header of each of the data files.

12. A data management method, comprising the steps of:

10 restoring a plurality of received data files, the data files being capable of being different file types;

organizing/categorizing the received data files, based on a predetermined user list, into a source directory structure and a destination directory structure;

15 logging the received data files into a database formed by the source and destination directory structures and identifying a file type of the received data files;

de-duplicating duplicates in the received data files by calculating a SHA value of the received data files to determine whether the received data files have duplicates and flagging the duplicated data files in the database;

20 converting the remaining data files into image files, respectively; and exporting the image files.

13. The method of claim 12, further comprising the step of viewing the image files stored in the database.

5 15. The method of claim 12, wherein the identifying of the data files includes identifying email data files and user data files, the email data files are in a variety of formats including Microsoft Mail, Outlook, GroupWise, Lotus Notes, the user data files have a variety of formats including Word, Excel, PowerPoint, and Access, the email data files include attachment data and email files.

10

16. The method of claim 15, further comprising the step of associating the email data files with the corresponding attachment data and email files such that the image data files for the email data files and the corresponding attachment data and email files can be viewed together.

15

17. The method of claim 12, further comprising the step of parallel processing the steps of logging, converting, and exporting such that the data files are parallel-processed in a data file logging stage, an image conversion stage, and an image file output stage.

20 18. The method of claim 12, wherein the converting of the data files includes
converting the data files having the same file type into the image files together.

19. The method of claim 12, wherein the converting of the data files is processed by a plurality of image conversion processors, each of the image conversion processors being capable of converting the data files having the same file type into the corresponding image files.

5

20. The method of claim 12, wherein the identifying of the file type of the data files is based on the SHA value and a file header of each of the data files.